Alea 0, 0–0 (2014)

Estimation for Stochastic Damping Hamiltonian Systems under partial observation. II. Drift term.

P. Cattiaux, J. León and C. Prieur

Institut de Mathématiques de Toulouse. Université de Toulouse.
CNRS UMR 5219.
118 route de Narbonne, 31062 Toulouse cedex 09, France.
E-mail address: `cattiaux@math.univ-toulouse.fr`

Université Joseph Fourier, LJK/MOISE BP 53.
38041 Grenoble Cedex, France.
E-mail address: `clementine.prieur@imag.fr`

Escuela de Matemática, Facultad de Ciencias, Universidad Central de Venezuela.
A.P. 47197, Los Chaguaramos, Caracas 1041-A, Venezuela.
E-mail address: `jose.leon@ciens.ucv.ve`

Abstract. This paper is the second part of our study started with Cattiaux et al. (2014). For some ergodic hamiltonian systems we obtained a central limit theorem for a non-parametric estimator of the invariant density, under partial observation (only the positions are observed). Here we obtain similarly a central limit theorem for a non-parametric estimator of the drift term. This theorem relies on the previous result for the invariant density.

1. INTRODUCTION.

Let $\big(Z_t := (X_t, Y_t) \in \mathbb{R}^{2d}\,,\; t \geq 0\big)$ be governed by the following Ito stochastic differential equation:

$$\begin{aligned} dX_t &= Y_t dt \\ dY_t &= \sigma\, dW_t - (c(X_t, Y_t)Y_t + \nabla V(X_t))dt. \end{aligned} \tag{1.1}$$

Each component Y^i $(1 \leq i \leq d)$ is the velocity of a particle i with position X^i. Function c is called the damping force and V the potential, σ is some (non-zero) constant and W a standard brownian motion.
We shall assume that c and V are regular enough for the existence and uniqueness of a non explosive solution of (1.1). We shall also assume that the process is ergodic

2010 *Mathematics Subject Classification.* 60H10; 62M05; 62G08; 35H10.

Key words and phrases. Hypoelliptic diffusion; Nonparametric drift estimation; Partial observations; High frequency estimation.

Research partially supported by the project ECOS NORD/FONACIT (France-Venezuela) **V12M01** *Equations différentielles stochastiques et analyse de sensibilité : applications environnementales.*

with a unique invariant probability measure μ, and that the convergence in the ergodic theorem is quick enough. Some sufficient conditions will be recalled in the next section.
In our previous work Cattiaux et al. (2014) we proposed a non-parametric estimator for the invariant density p_s of the invariant measure μ. We refer to the introduction of Cattiaux et al. (2014) for some references on this problem, as well as short discussion of the physical interest of such models.
In the present paper we attack the problem of estimating the drift term

$$g(x,y) = -\left(c(x,y)y + \nabla V(x)\right) \tag{1.2}$$

for incomplete but high-frequency data. We indeed observe only the first component of $(Z_t\,,\ t \geq 0)$, and our asymptotic results are proved under the assumption that this component is available at arbitrarily short inter-observation time intervals.

As explained in Cattiaux et al. (2014), while there is an impressive literature on non-parametric estimation for the invariant density or the drift term, most of it deals with elliptic diffusion processes. Here we are looking at a fully degenerate process, but still hypo-elliptic. In addition we intend to propose an estimator based on the observation of the positions X only, at some discretized observation times. Actually some works have already been done, in the hypo-elliptic context, but in a parametric framework. We refer to the work by Pokern et al. (2009) or to the recent work by Samson and Thieullen (2012) and the bibliography therein. In this paper we focus on pointwise estimation, as it is a not straightforward issue in itself. There exist results for the non-parametric estimation of the drift of one-dimensional ergodic diffusions, see e.g. Comte et al. (2007) in the context of discretized observations, or Dalalyan (2005) in the context of full time-continuous observations. For partial observations, there are results in the parametric setting (see e.g. Gloter, 2007). A natural development of our paper would be to extend this kind of results to our framework.

The main result of Cattiaux et al. (2014) reads as follows: if p_s denotes the invariant density (see the next section for its existence), then one can find a discretization step h_n, bandwidths b_{1n} and b_{2n} and kernels K such that, defining the estimator

$$\hat{p}_n(x,y) := \frac{1}{n b_{1n}^d b_{2n}^d} \sum_{i=1}^{n} K\left(\frac{x - X_{ih_n}}{b_{1n}}, \frac{y - \frac{X_{(i+1)h_n} - X_{ih_n}}{h_n}}{b_{2n}}\right),$$

corresponding to partial observation, it holds

$$\sqrt{n b_{1n}^d b_{2n}^d}\,\left(\hat{p}_n(x,y) - p_s(x,y)\right) \xrightarrow[n\to+\infty]{\mathcal{D}} \mathcal{N}\left(0, p_s(x,y) \int K^2(u,v)dudv\right),$$

for all pairs (x,y). The previous convergence in distribution holds true under the stationary distribution. In the non stationary case we have to shift the summation. See Theorem 3.1 and the comment following the statement of the Theorem in section 3.

The intuitive explanation for the definition of our estimator is based on the Nadaraya-Watson kernel method for regression estimation. Assuming that we observe both coordinates $(X_t, Y_t)_{t\geq 0}$ on a grid ih_n, $i = 1, \ldots, n$, we deduce from the system (1.1) the following approximation

$$Y_{(i+1)h_n} - Y_{ih_n} \approx \sigma(W_{(i+1)h_n} - W_{ih_n}) + g(X_{ih_n}, Y_{ih_n})h_n.$$

The first term is interpreted as the noise part and the second one as the regression term. The Naradaya-Watson's method allows us to introduce the following estimator of the function $g(x,y)p_s(x,y)$:

$$\breve{g}_n(x,y)\hat{p}_n(x,y) \quad = \quad \frac{1}{(n-1)b_{1n}b_{2n}} \sum_{i=1}^{n-1} K(\frac{x-X_{ih_n}}{b_1 n}, \frac{y-Y_{ih_n}}{b_2 n}) \frac{Y_{(i+1)h_n}-Y_{ih_n}}{h_n^{1/2}}$$

$$= \frac{1}{(n-1)b_{1n}b_{2n}} \sum_{i=1}^{n-1} K(\frac{x-X_{ih_n}}{b_1 n}, \frac{y-Y_{ih_n}}{b_2 n}) \left(\frac{\sigma(W_{(i+1)h_n}-W_{ih_n})}{h_n^{1/2}} + g(X_{ih_n}, Y_{ih_n})h_n^{1/2} \right).$$

We then have

$$\mathbb{E}(\frac{\breve{g}_n(x,y)\hat{p}_n(x,y)}{h_n^{1/2}}) \to g(x,y)p_s(x,y).$$

The estimator $\breve{g}_n(x,y)$ is an thus an asymptotically unbiased estimator of $g(x,y)$.

In our work we assume that only the first coordinate X_t is observed. Thus we define an estimator where

$$Y_{ih_n} \text{ is replaced by } X_{(i+1)h_n} - X_{ih_n}$$

and

$$Y_{(i+1)h_n} - Y_{ih_n} \text{ by } X_{(i+1)h_n} - 2X_{ih_n} + X_{(i-1)h_n}.$$

Furthermore, to preserve some independence properties we need to have a certain lag between the observations used in the kernel and those defining the double increment. This last observation leads us to define $X_{(i+\frac{1}{3})h_n} - X_{ih_n}$ for the first increment and $X_{(i+1)h_n} - 2X_{(i+\frac{2}{3})h_n} + X_{(i+\frac{1}{3})h_n}$ for the second one.

We thus define the estimator $\hat{g}_n$ as

$$\hat{g}_n(x,y)\,\hat{p}_n(x,y) := \frac{1}{(n-1)b_{1n}^d b_{2n}^d} \sum_{i=1}^{n-1} K\left(\frac{x-X_{ih_n}}{b_{1n}}, \frac{y - \frac{X_{(i+\frac{1}{3})h_n}-X_{ih_n}}{(h_n/3)}}{b_{2n}} \right) \frac{\mathfrak{D}_{i,n}}{(h_n/3)^2},$$

where

$$\mathfrak{D}_{i,n} := X_{(i+1)h_n} - 2X_{(i+\frac{2}{3})h_n} + X_{(i+\frac{1}{3})h_n}.$$

Define $\hat{H}_n(x,y) := \hat{g}_n(x,y)\,\hat{p}_n(x,y)$. In Theorem 4.1 of the present paper we state that one can find h_n and two different bandwidths b_{in} (for the definition of $\hat{H}_n$) and c_{in} (for the definition of $\hat{p}_n$) $(i=1,2)$ such that in the stationary regime

$$\sqrt{nb_{1n}^d b_{2n}^d h_n}\,(\hat{g}_n(x,y) - g(x,y)) \xrightarrow[n\to+\infty]{\mathcal{D}} \mathcal{N}\left(0, (\frac{2}{3}\frac{\sigma^2}{p_s(x,y)} \int K^2(u,v)dudv)\mathcal{I}\right),$$

where $\mathcal{I}$ denotes the identity matrix in $\mathbb{R}^d$. As for the invariant density we also prove such a central limit theorem in the non-stationary regime, shifting the summation (see section 5).

It should be very interesting to estimate separately $c(x,y)$ and $\nabla V(x)$, who have different physical interpretations. Actually, there is no explicit relation between the invariant density and the drift term, unless c is constant, in which case $p_s(x,y) = \exp\left(-\frac{2c}{\sigma^2}\left(\frac{|y|^2}{2} + V(x)\right)\right)$. In that particular case, the estimation of the potential V and of its gradient ∇V can be deduced from the estimators of the invariant density and of the drift term. In full generality this will require some other ideas.

2. THE MODEL AND ITS PROPERTIES.

We are obliged to recall some facts on the model. A more detailed discussion is contained in Cattiaux et al. (2014).

We shall first give some results about non explosion and long time behaviour. In a sense, coercivity can be seen in this context as some exponential decay to equilibrium.

Let us first introduce some sets of assumptions:

Hypothesis $\mathcal{H}_1$:

(i) the potential V is lower bounded, smooth over $\mathbb{R}^d$, V and ∇V have polynomial growth at infinity and there exists $v > 0$ such that

$$+\infty \geq \liminf_{|x|\to+\infty} \frac{x.\nabla V(x)}{|x|} \geq v > 0\,,$$

the latter being often called "drift condition",

(ii) the damping coefficient $c(x,y)$ is smooth and bounded, and there exist $c, L > 0$ so that $c^s(x,y) \geq cId > 0$, $\forall(|x| > L, y \in \mathbb{R}^d)$, where $c^s(x,y)$ is the symmetrization of the matrix $c(x,y)$, given by $\frac{1}{2}(c_{ij}(x,y)+c_{ji}(x,y))_{1\leq i,j\leq d}$,

These conditions ensure that there is no explosion, and that the process is positive recurrent with a unique invariant probability measure μ. We will denote by $P_t f(z) = \mathbb{E}_z(f(Z_t))$ which is well defined for all bounded function f, P_t extends as a contraction semi-group on $\mathbb{L}^p(\mu)$ for all $1 \leq p \leq +\infty$. Further in the paper we add assumptions to ensure that the drift function g defined by (1.2) belongs to the domain of the infinitesimal generator L of P_t (see Assumptions $\mathcal{H}_3$ and $\mathcal{H}_4$).

Furthermore μ admits some exponential moment, hence polynomial moments of any order. Another key feature is that the process is actually α-mixing, i.e.

Proposition 2.1. *There exist some constants $C > 0$ and $\rho < 1$ such that:*

$$\forall\, g, f \in \mathbb{L}^\infty(\mu)\,, \quad \forall\, t \geq 0,$$
$$|\mathrm{Cov}_\mu\,(f(Z_t), g(Z_0))| \leq C\,\rho^{t/2}\,\left\|g - \int g d\mu\right\|_\infty \left\|f - \int f d\mu\right\|_\infty . \tag{2.1}$$

i.e., in the stationary regime, $(Z_t,\, t \geq 0)$ is α-mixing with exponential rate.

As explained in section 2.2 of Cattiaux et al. (2014), the infinitesimal generator L is hypo-elliptic, which implies that

$$\mu(dz) = p_s(z)\,dz$$

with some smooth function p_s. One can relax the C^∞ assumption on the coefficients into a C^k assumption, for a large enough k, but this is irrelevant.

Furthermore it can be shown that p_s is everywhere positive, for instance by using an extension of Girsanov theory which is available here.

One can relax some assumptions and still have the same conclusions:

Hypothesis $\mathcal{H}_2$:

(a) One can relax the boundedness assumption on c in $\mathcal{H}_1$, assuming that for all $N > 0$: $\sup_{|x|\leq N, y\in\mathbb{R}^d} \|c(x,y)\|_{H.S.} < +\infty$, where $H.S.$ denotes the Hilbert-Schmidt norm of a matrix; but one has to assume in addition conditions (3.1) and (3.2) in Wu (2001). An interesting example (the Van der Pol model) in this situation is described in Wu (2001) subsection 5.3.

(b) The most studied situation is the one when c is a constant matrix. Actually almost all results obtained in Wu (2001) or Bakry et al. (2008) in this situation extend to the general bounded case.
Nevertheless we shall assume now that c is a constant matrix.
In this case a very general statement replacing $\mathcal{H}_1$ (i) is given in Theorem 6.5 of Bakry et al. (2008). Tractable examples are discussed in Example 6.6 of the same paper. In particular one can replace the drift condition on V by

$$\liminf_{|x|\to+\infty} |\nabla V|^2(x) > 0 \quad \text{and} \quad \|\nabla^2 V\|_{H.S.} \ll |\nabla V| \, .$$

Notice that one can relax the repelling strength of the potential, and obtain, no more exponential but sub-exponential or polynomial decay (see the discussion in Bakry et al. (2008)).

From now on in the whole paper we will assume that Hypothesis $\mathcal{H}_1$ (or $\mathcal{H}_2$) is fulfilled.
In all the proofs of the paper C denotes some constant which may vary from line to line.

3. Estimation of the invariant density.

In this section we recall the central limit theorem for a non-parametric estimator of the invariant density p_s proposed in Cattiaux et al. (2014).

First we consider that one can observe the whole process $Z_.$ at discrete times with discretization step h_n, i.e we consider

$$\tilde{p}_n(x,y) := \frac{1}{n b_{1n}^d b_{2n}^d} \sum_{i=1}^{n} K\left(\frac{x - X_{ih_n}}{b_{1n}}, \frac{y - Y_{ih_n}}{b_{2n}}\right) . \tag{3.1}$$

Second we consider the partially observed case, where only the position process $X_.$ can be observed, and we approximate the velocity, i.e. we consider

$$\hat{p}_n(x,y) := \frac{1}{n b_{1n}^d b_{2n}^d} \sum_{i=1}^{n} K\left(\frac{x - X_{ih_n}}{b_{1n}}, \frac{y - \frac{X_{(i+1)h_n} - X_{ih_n}}{h_n}}{b_{2n}}\right) . \tag{3.2}$$

In both cases, the kernel K is some C^2 function with compact support A such that $\int_A K(x,y)dxdy = 1$. We may also assume, without loss of generality that A is a bounded ball. Moreover, we assume that there exists $m \in \mathbb{N}^*$ such that for all polynomials $P(x,y)$ with degree between 1 and m, $\int P(u,v)K(u,v)dudv = 0$. That is, we assume the kernel K is of order m.

Let us state the main result in this section (Theorem 3.1 below).

Theorem 3.1 (Cattiaux et al. (2014))**.** *Assume Hypothesis $\mathcal{H}_1$ or $\mathcal{H}_2$ are fulfilled. Recall that p_s denotes the density of the invariant measure μ. Assume that the* *bandwidths b_{1n}, b_{2n} and the discretization step h_n tend to zero as n tends to infinity and satisfy the following assumptions:*

(i) $n\, b_{1n}^d\, b_{2n}^d \to +\infty$,

(ii) $\frac{b_{1n}\, b_{2n}}{h_n^2} \to 0\,,$

(iii) *m is such that* $n\, b_{1n}^d\, b_{2n}^d\, \max(b_{1n}, b_{2n})^{2(m+1)} \to 0$.

Then, in the stationary regime, one gets for any $(x, y) \in \mathbb{R}^{2d}$

$$\sqrt{nb_{1n}^d b_{2n}^d}\,(\tilde{p}_n(x,y) - p_s(x,y)) \xrightarrow[n\to+\infty]{\mathcal{D}} \mathcal{N}\left(0, p_s(x,y)\int K^2(u,v)dudv\right)\,.$$

If in addition

(iv) $nh_n\, \frac{b_{1n}^d}{b_{2n}^{2+d}} \to 0$,

(v) *there exists $p > 1$ such that* $n\, h_n^2\, \frac{b_{1n}^{d(2-p)/p}}{b_{2n}^{2+d}} \to 0$.

Then, still in the stationary regime, one gets for any $(x, y) \in \mathbb{R}^{2d}$

$$\sqrt{nb_{1n}^d b_{2n}^d}\,(\hat{p}_n(x,y) - p_s(x,y)) \xrightarrow[n\to+\infty]{\mathcal{D}} \mathcal{N}\left(0, p_s(x,y)\int K^2(u,v)dudv\right)\,.$$

A similar statement holds true starting from any point $z_0 \in \mathbb{R}^{2d}$. In this situation we have to slightly change the definition of our estimators replacing $\sum_{i=1}^{n}$ by $\sum_{i=1+l_n}^{n+l_n}$ for some l_n such that $l_n\, h_n \to +\infty$ as $n \to +\infty$.

4. Estimation of the drift term.

In this section we will consider an estimator of the drift function from $\mathbb{R}^{2d}$ into $\mathbb{R}^d$, $g(x, y) = -[c(x, y)y + \nabla V(x)]$.

Let $K : \mathbb{R}^{2d} \to \mathbb{R}$ be a C^2 $2d$-dimensional kernel whose support is compact and such that there exists $m \in \mathbb{N}^*$ such that for all non constant polynomial $P(x, y)$ with degree less or equal than m, $\int P(u, v)K(u, v)dudv = 0$. The estimators of the invariant density, $\tilde{p}_n(x, y)$ and $\hat{p}_n(x, y)$ are defined as in Section 3. However, for simplicity we will only use $\hat{p}_n(x, y)$.

Define for $i \in \mathbb{N}^*$, $n \in \mathbb{N}^*$,

$$\begin{aligned} \mathfrak{D}_{i,n} &:= X_{(i+1)h_n} - 2X_{(i+\frac{2}{3})h_n} + X_{(i+\frac{1}{3})h_n} \\ &= \int_{(i+\frac{2}{3})h_n}^{(i+1)h_n} (Y_s - Y_{(i+\frac{2}{3})h_n})\, ds + \int_{(i+\frac{1}{3})h_n}^{(i+\frac{2}{3})h_n} (Y_{(i+\frac{2}{3})h_n} - Y_s)\, ds\,. \end{aligned} \tag{4.1}$$

Because one observes only the position of the particle and not its derivative, we must define our estimator by using only the position.

We introduce the estimator $\hat{g}_n(x, y)$ of $g(x, y)$ defined as

$$\hat{g}_n(x,y)\,\hat{p}_n(x,y) := \frac{1}{(n-1)b_{1n}^d b_{2n}^d}\sum_{i=1}^{n-1} K\left(\frac{x - X_{ih_n}}{b_{1n}}, \frac{y - \frac{X_{(i+\frac{1}{3})h_n} - X_{ih_n}}{(h_n/3)}}{b_{2n}}\right)\frac{\mathfrak{D}_{i,n}}{(h_n/3)^2}\,,$$

where

$$\hat{p}_n(x,y) := \frac{1}{nb_{1n}^d b_{2n}^d} \sum_{i=1}^{n} K\left(\frac{x - X_{ih_n}}{c_{1n}}, \frac{y - \frac{X_{(i+1)h_n} - X_{ih_n}}{h_n}}{c_{2n}}\right).$$

Note that the the bandwidths b_{in}, $i = 1,2$ and c_{in}, $i = 1,2$ are constrained by (4.2) in Theorem 4.1. This constraint is needed for our proof of Theorem 4.1 from Proposition 4.3 which deals with the central limit theorem for $\hat{H}_n$. The choice of our estimator $\hat{g}_n$ has been motivated in the introduction of the paper. It is both based on a natural heuristic and on technical arguments needed for the proof of Proposition 4.2, which is a step towards the proof of our main result. This main result is stated in Theorem 4.1 below, it consists in a central limit theorem for $\hat{g}_n(x,y)$.

During the proof we shall need an additional assumption namely:

Hypothesis $\mathcal{H}_3$: g belongs to the domain of the infinitesimal generator L, in all $\mathbb{L}^p(\mu)$ for $1 \leq p < +\infty$.

According to the properties we recalled before, for $\mathcal{H}_3$ to be satisfied it is enough that:

Hypothesis $\mathcal{H}_4$: the function c (resp. V) and its first two derivatives (resp. its first three derivatives) have polynomial growth.

The aim of these assumptions is to ensure that the drift function g belongs to the domain of the infinitesimal generator L of P_t.

Theorem 4.1. *Assume that $\mathcal{H}_1$, $\mathcal{H}_2$ and $\mathcal{H}_3$ are satisfied and that the the bandwidths b_{1n}, b_{2n}, c_{1n}, c_{2n} and the discretization step h_n tend to zero as n tends to ∞. Assume moreover that the following assumptions are satisfied respectively with $r_{in} = b_{in}$ and $r_{in} = c_{in}$:*

i) $n\, h_n\, r_{1n}^d\, r_{2n}^d \to +\infty$,

ii) *$m \in \mathbb{N}^*$ is such that* $nr_{1n}^d r_{2n}^d h_n \max(r_{1n}, r_{2n})^{2(m+1)} \xrightarrow[n\to+\infty]{} 0$,

iii) *$\exists\, \varepsilon_1 > 0$ such that* $n(r_{1n}^d r_{2n}^d)^{1-\varepsilon_1}\, h_n^3 \xrightarrow[n\to+\infty]{} 0$,

iv) *$\exists\, \varepsilon_2, \varepsilon_3 \in \mathbb{R}_+^*$, $\varepsilon_2, \varepsilon_3 < 1$ such that* $\frac{h_n^{2(1-\varepsilon_2)}}{(r_{1n} r_{2n})^{\varepsilon_3}} \xrightarrow[n\to+\infty]{} 0$,

v) *there exist $p > 1$, $p < +\infty$ and $\varepsilon > 0$, such that* $h_n^2 r_{1n}^{d(\frac{1}{p}-1)} r_{2n}^{-(2+d)} \xrightarrow[n\to+\infty]{} 0$ *and* $h_n \sqrt{n}\, r_{1n}^{d(\frac{1}{p(1+\varepsilon)}-\frac{1}{2})} r_{2n}^{-(\frac{d}{2}+1)} \xrightarrow[n\to+\infty]{} 0$.

Define $\hat{g}_n = \hat{H}_n(b)/\hat{p}_n(c)$ where the indication in brackets indicates the bandwidths we are using. Assume in addition that

$$h_n(b,c) := h_n \frac{b_{1n}^d b_{2n}^d}{c_{1n}^d c_{2n}^d} \to 0 \quad \text{as } n \to +\infty. \tag{4.2}$$

Then

$$\sqrt{nb_{1n}^d b_{2n}^d h_n}\, (\hat{g}_n(x,y) - g(x,y)) \xrightarrow[n\to+\infty]{\mathcal{D}} \mathcal{N}\left(0, (\frac{2}{3}\frac{\sigma^2}{p_s(x,y)} \int K^2(u,v) du dv)\mathcal{I}\right).$$

For technical reasons, the proof of Theorem 4.1 is based on two preliminary results stated respectively in Proposition 4.2 and Proposition 4.3. We first introduce an intermediate kernel estimate of $g(x,y)\,p_s(x,y)$, denoted by $\tilde{H}_n$, and defined by:

$$\tilde{H}_n(x,y) = \frac{1}{(n-1)b_{1n}^d b_{2n}^d} \sum_{i=1}^{n-1} K\left(\frac{x - X_{ih_n}}{b_{1n}}, \frac{y - Y_{ih_n}}{b_{2n}}\right) \frac{\mathfrak{D}_{i,n}}{(h_n/3)^2}\,. \tag{4.3}$$

Note that this estimate can not be the final estimate as the Y_{ih_n}s are not observed in practice. However, studying the convergence of $\tilde{H}_n(x,y)$ is easier and is the first setp towards the proof of Theorem 4.1.

Now, to see why the definition of $\tilde{H}_n(x,y)$ is meaningful, let us first study the asymptotic bias of this estimator in the stationary regime.

Using stationarity we get

$$\begin{aligned}
\mathbb{E}[\tilde{H}_n(x,y)] &= \frac{9}{b_{1n}^d b_{2n}^d}\mathbb{E}\left[K\left(\frac{x - X_0}{b_{1n}}, \frac{y - Y_0}{b_{2n}}\right)\frac{\mathfrak{D}_{0,n}}{h_n^2}\right] \\
&= \frac{9}{b_{1n}^d b_{2n}^d}\int_{\mathbb{R}^{2d}} K\left(\frac{x-u}{b_{1n}}, \frac{y-v}{b_{2n}}\right)\mathbb{E}\left[\frac{\mathfrak{D}_{0,n}}{h_n^2}|X_0 = u\, Y_0 = v\right] p_s(u,v)dudv.
\end{aligned}$$

Using (4.1), we may write

$$\frac{\mathfrak{D}_{0,n}}{h_n^2} = \frac{1}{h_n^2}\big(\sigma[\int_{\frac{2}{3}h_n}^{h_n}(W_s - W_{\frac{2}{3}h_n})ds + \int_{\frac{h_n}{3}}^{\frac{2}{3}h_n}(W_{\frac{2}{3}h_n} - W_s)ds] + I(h_n)\big), \tag{4.4}$$

where

$$I(h_n) = \int_{\frac{2}{3}h_n}^{h_n}\int_{\frac{2}{3}h_n}^{t} g(X_s,Y_s)dsdt + \int_{\frac{h_n}{3}}^{\frac{2}{3}h_n}\int_{t}^{\frac{2}{3}h_n} g(X_s,Y_s)dsdt.$$

The independence of the increments of W and the semigroup properties yield

$$\begin{aligned}
\mathbb{E}[\frac{\mathfrak{D}_{0,n}}{h_n^2}|X_0 = u\,,\, Y_0 = v] &= \frac{1}{h_n^2}\left[\int_{\frac{2}{3}h_n}^{h_n}\int_{\frac{2}{3}h_n}^{t} P_s g(u,v)dsdt + \int_{\frac{h_n}{3}}^{\frac{2}{3}h_n}\int_{t}^{\frac{2}{3}h_n} P_s g(u,v)dsdt\right] \\
&=: G(h_n,u,v).
\end{aligned}$$

Notice that, $G(h_n,u,v) \to \frac{1}{9}\,g(u,v)$ as $n \to +\infty$ since $h_n \to 0$.
A change of variable entails

$$\begin{aligned}
\mathbb{E}[\tilde{H}_n(x,y)] &= \frac{9}{b_{1n}^d b_{2n}^d}\int_{\mathbb{R}^{2d}} K\left(\frac{x-u}{b_{1n}}, \frac{y-v}{b_{2n}}\right) G(h_n,u,v)p_s(u,v)dudv \\
&= 9\int_{\mathbb{R}^{2d}} K(z_1,z_2)G(h_n, x - b_{1n}z_1, y - b_{2n}z_2)p_s(x - b_{1n}z_1, y - b_{2n}z_2)dz_1dz_2\,,
\end{aligned}$$

which converges to $g(x,y)p_s(x,y)$ as n tends to infinity, according to the bounded convergence theorem and under Assumption $\mathcal{H}_4$. Indeed, Assumption $\mathcal{H}_4$ ensures that g defined by (1.2) is regular enough to belong to the domain of the infinitesimal generator of the semi-group P_t. The main argument to prove that is given in Lemma 1.1 in Wu (2001). Hence $\tilde{H}_n$ is an asymptotically unbiased estimator of gp_s.

Starting from this consideration, we now state a central limit theorem for the estimator $\tilde{H}_n(x,y)$. Let denote $\mathcal{I}$ the identity matrix in $\mathbb{R}^d$.

Proposition 4.2. *Assume that $\mathcal{H}_1$, $\mathcal{H}_2$ and $\mathcal{H}_3$ are satisfied and that the bandwidths b_{1n}, b_{2n} and the discretization step h_n tend to zero as n tends to ∞. Assume moreover that the bandwidths b_{in}, $i = 1, 2$, satisfy assumptions i) to iv) of Theorem 4.1. Then, in the stationary regime,*

$$\sqrt{nb_{1n}^d b_{2n}^d h_n}\left(\tilde{H}_n(x,y) - g(x,y)p_s(x,y)\right) \xrightarrow[n\to+\infty]{\mathcal{D}} \mathcal{N}\left(0, (\frac{2}{3}\sigma^2 p_s(x,y)\int K^2(u,v)dudv)\mathcal{I}\right).$$

Proof of Proposition 4.2: the proof is postponed to the Appendix. □

We now deduce from Proposition 4.2 the following result on the asymptotic behaviour of $\hat{H}_n(x,y) = \hat{g}_n(x,y)\,\hat{p}_n(x,y)$:

Proposition 4.3. *Under the assumptions of Proposition 4.2 and assuming moreover assumption v) of Theorem 4.1 one gets*

$$\sqrt{nb_{1n}^d b_{2n}^d h_n}\left(\hat{H}_n(x,y) - p_s(x,y)g(x,y)\right) \xrightarrow[n\to+\infty]{\mathcal{D}} \mathcal{N}\left(0, (\frac{2}{3}\sigma^2 p_s(x,y)\int K^2(u,v)dudv)\mathcal{I}\right).$$

Proof of Proposition 4.3: the proof is postponed to the Appendix. □

Proof of the main result, Theorem 4.1: We have

$$\hat{g}_n - g = \frac{\hat{H}_n}{\hat{p}_n} - g = \frac{\hat{H}_n - p_s g}{p_s} + \hat{H}_n\left(\frac{1}{\hat{p}_n} - \frac{1}{p_s}\right).$$

For the first term, we have according to Proposition 4.3,

$$\sqrt{nb_{1n}^d b_{2n}^d h_n}\,\frac{\hat{H}_n - p_s g}{p_s}(x,y) \xrightarrow[n\to+\infty]{\mathcal{D}} \mathcal{N}\left(0, (\frac{2}{3}\frac{\sigma^2}{p_s(x,y)}\int K^2(u,v)dudv)\mathcal{I}\right).$$

We shall show that the second one goes to 0 in probability and then conclude by using Slutsky theorem.
We thus decompose

$$\sqrt{nb_{1n}^d b_{2n}^d h_n}\,\hat{H}_n\left(\frac{1}{\hat{p}_n} - \frac{1}{p_s}\right) = \left(\frac{\sqrt{h_n(b,c)}\,\hat{H}_n}{\hat{p}_n\,p_s}\right)\left(\sqrt{nc_{1n}^d c_{2n}^d}\,(p_s - \hat{p}_n)\right).$$

The second term of this product converges in distribution according to Theorem 3.1. We shall show that the first term in the product goes to 0 in probability.
Indeed, according to Cattiaux et al. (2014), $\hat{p}_n - p_s \to 0$ in probability as $n \to +\infty$, and as we previously saw, $\hat{H}_n$ is bounded in $\mathbb{L}^1$. Let $a > 0$. We have

$$\mathbb{P}\left(\frac{\sqrt{h_n(b,c)}\,\hat{H}_n}{\hat{p}_n\,p_s} > a\right) \leq \mathbb{P}\left(\hat{p}_n \leq (p_s/2)\right) + \mathbb{P}\left(\hat{H}_n > (a\,p_s^2/2\sqrt{h_n(b,c)})\right),$$

and both terms go to 0, using the Markov inequality for instance for the second one.
To conclude it remains to recall that if U_n goes to 0 in probability and V_n goes to V in distribution, the product $U_n\,V_n$ goes to 0 in probability. □.

We conclude this section by giving an explicit class of examples for the parameters h_n, $b_{i,n}$, $c_{i,n}$ to satisfy all the required assumptions in Theorem 4.1.

Proposition 4.4. *Choose $h_n = n^{-\gamma}$, $b_{in} = n^{-\alpha_i}$ and $c_{in} = n^{-\beta_i}$ for some $\gamma, \alpha_i, \beta_i > 0$. If*

(1) $\alpha_2 = \beta_2 - \frac{\varepsilon}{d(4+4d)} = \frac{1-2\varepsilon}{d(4+4d)}$,
(2) $\beta_1 = \frac{3-2\varepsilon+4d}{d(4+4d)}$, $\alpha_1 = \beta_1 + \frac{\beta_2}{2} - \frac{1}{2d^2}$,
(3) $\frac{1}{2d} < \gamma < \frac{1}{2d} + \frac{3\varepsilon}{4+4d}$,
(4) $m > \frac{1+2d}{1-\varepsilon}$.

for some $\varepsilon > 0$ *small enough, then Theorem 4.1 holds true.*

Remark 4.5. Under assumptions of Proposition 4.4 we can reach a rate of convergence in $n^{\frac{3\varepsilon}{4+4d}}$.

Proof of Proposition 4.4: the proof is postponed to the appendix. □

Remark 4.6. Remark that contrary to the case of Theorem 3.1, where it is possible to have $m=1$ here m grows linearly with the dimension ($m > \frac{2+4d+6\varepsilon d+2\varepsilon}{2(1-\varepsilon)}$).

5. Non-stationary case

In Section 3 we stated the central limit theorem for the estimate of the drift $g(x,y)$ in the case where the process is in the stationary regime. Let us now define the new estimate

$$\overline{g}_n(x,y) = \frac{\frac{1}{(n-1)b_{1n}^d b_{2n}^d} \sum_{i=l_n+1}^{n+l_n-1} K\left(\frac{x-X_{ih_n}}{b_{1n}}, \frac{y-\frac{X_{(i+\frac{1}{3})h_n}-X_{ih_n}}{(h_n/3)}}{b_{2n}}\right)\frac{\mathfrak{D}_{i,n}}{(h_n/3)^2}}{\frac{1}{nb_{1n}^d b_{2n}^d} \sum_{i=l_n+1}^{n+l_n} K\left(\frac{x-X_{ih_n}}{b_{1n}}, \frac{y-\frac{X_{(i+1)h_n}-X_{ih_n}}{h_n}}{b_{2n}}\right)} \tag{5.1}$$

We remark that given $Z_0 \sim \mu(dz)$, $\overline{g}_n(x,y) \overset{\mathcal{L}}{=} \hat{g}_n(x,y)\ \forall\, n \in \mathbb{N}^*$.

Theorem 5.1 below states that we can estimate $g(x,y)$ by using $\overline{g}_n(x,y)$ with $Z_0 = z_0 = (x_0, y_0)$.

Theorem 5.1. *Under the assumptions in Theorem 4.1, starting from any initial point* $z_0 = (x_0, y_0)$*, it holds*

$$\sqrt{n\, h_n\, b_{1n}^d b_{2n}^d}\,(\overline{g}_n(x,y) - g(x,y)) \xrightarrow[n\to+\infty]{\mathcal{D}} \mathcal{N}\left(0, (\frac{2}{3}\frac{\sigma^2}{p_s(x,y)}\int K^2(u,v)dudv)\mathcal{I}\right),$$

provided l_n *appearing in the definition (5.1) of* $\overline{g}_n(x,y)$ *satisfies* $l_n h_n \xrightarrow[n\to+\infty]{} +\infty$.

Proof of Theorem 5.1:
Recall that $p_s : \mathbb{R}^{2d} \mapsto \mathbb{R}_+$ denotes the invariant density of Z and μ the associated invariant probability measure.

Denote by $\mathcal{C}_b(\mathbb{R})$ the set of bounded continuous functions $h : \mathbb{R} \to \mathbb{R}$. It is only necessary to prove that, for any $h \in \mathcal{C}_b(\mathbb{R})$, the difference

$$\begin{aligned}\Delta_n(h) &= \mathbb{E}\Big[h(\sqrt{h_n\, nb_{1n}^d b_{2n}^d}\hat{g}_n(x,y)|Z_0\sim\mu) - h(\sqrt{h_n\, nb_{1n}^d b_{2n}^d}\overline{g}_n(x,y)|Z_0 = z_0)\Big]\\ &= \mathbb{E}\Big[h(\sqrt{h_n\, nb_{1n}^d b_{2n}^d}\overline{g}_n(x,y)|Z_0\sim\mu) - h(\sqrt{h_n\, nb_{1n}^d b_{2n}^d}\overline{g}_n(x,y)|Z_0 = z_0)\Big]\end{aligned}$$

goes to zero as n tends to infinity. Let $h \in \mathcal{C}_b(\mathbb{R})$ and denote $\theta = \|h\|_\infty$.

To evaluate $\mathbb{E}\,(h(\overline{g}_n(x,y))|Z_0\sim\mu) - \mathbb{E}\,(h(\overline{g}_n(x,y))|Z_0 = z_0)$. Let us fix $n \in \mathbb{N}^*$. We first make the computations conditionally to Z_{jh_n}, $j > l_n + 1$.

One may write
$\overline{g}_n(x,y) = \overline{g}_n(x,y,Z_{(l_n+1)h_n}, Z_{jh_n}, j > l_n+1)$ and
$h_{Z_{jh_n}, j>l_n+1}(z') = \overline{g}_n(x,y,z', Z_{jh_n}, j > l_n+1)$.

Now, conditionally to Z_{jh_n}, $j > l_n+1$, one has:

$$
\begin{aligned}
&|\mathbb{E}\left(h(\overline{g}_n(x,y))|Z_0 \sim \mu\right) - \mathbb{E}\left(h(\overline{g}_n(x,y))|Z_0 = z_0\right)| \\
&= \left| \int h_{Z_{jh_n}, j>l_n+1}(z') \left(p_s(z') - q_{(l_n+1)h_n}(z_0,z')\right) dz' \right| \\
&\leq \theta D \rho^{(l_n+1)h_n} \Psi(z_0)
\end{aligned}
\tag{5.2}
$$

using Inequality (2.1) in Cattiaux et al. (2014).

Finally, as $0 < \rho < 1$, we can conclude that $\Delta_n(h)$ goes to zero as n tends to infinity as soon as $l_n h_n \xrightarrow[n\to+\infty]{} +\infty$, which concludes the proof of Theorem 5.1. □

6. Examples and numerical simulation results

In this section, we consider examples of stochastic differential equations defined by (1.1) and implement the estimator on simulated data. However, the choice of the optimal bandwidths b_{1n}, b_{2n}, c_{1n} and c_{2n} as far as the choice of the optimal discretization step h_n, and the optimal choice of the kernel K, although interesting, is not the purpose of this section nor this paper. This is a separate study to be addressed in future work.

To simulate sample paths, we use an approximate discrete sampling generated by an explicit Euler scheme (see Remark 6.1 at the end of this section). We consider three specific examples. The first one has been proposed in Pokern et al. (2009). It corresponds to a linear oscillator subject to noise and damping. The second example is one example of generalized Duffing oscillators described in Wu (2001) subsection 5.2. The last example is the Van der Pol oscillator whose damping force depends on both position and velocity coordinates. These three models are of type (1.1) and satisfy assumptions needed to apply our estimation results. Simulations are run with the Epanechnikov kernel.

6.1. *Model I: harmonic oscillator.*

We consider an harmonic oscillator that is driven by a white noise forcing:

$$
\begin{aligned}
dX_t &= Y_t dt \\
dY_t &= \sigma\, dW_t - (\kappa Y_t + D X_t)dt.
\end{aligned}
\tag{6.1}
$$

with $\kappa > 0$ and $D > 0$. In the following we choose $D = 2$, $\kappa = 2$ and $\sigma = 1$. For this model we know that the stationary distribution is Gaussian, with mean zero and an explicit variance matrix given in e.g. Gardiner (1985). With our choice of parameters, the Gaussian invariant density is

$$
p_s(x,y) = \frac{2\sqrt{2}}{\pi} \exp\left(-4x^2 - 2y^2\right).
$$

And the drift is defined by $g(x,y) = -2(y+x)$. In the following we make use of the explicit Euler scheme to simulate an approximated discrete sampling $(\widetilde{X}_i, \widetilde{Y}_i)_{i\in\mathbb{N}}$ of $(X_t, Y_t)_{t\in\mathbb{R}_+}$. For a given step $\delta > 0$, the scheme is defined as

$$\begin{aligned} \widetilde{X}_{i+1} - \widetilde{X}_i &= \widetilde{Y}_i \delta \\ \widetilde{Y}_{(i+1)} - \widetilde{Y}_i &= \sigma\,(W_{(i+1)\delta} - W_{i\delta}) - (\kappa \widetilde{Y}_i + D\widetilde{X}_i)\delta \end{aligned} \tag{6.2}$$

$(\widetilde{X}_0, \widetilde{Y}_0) = (0,0)$. We take $n = 5000$, $h = 0.28$, $b_{1n} = b_{2n} = 0.18$, and the step for the explicit Euler scheme $\delta = \frac{1}{10}h/3$.

For some fixed value of x_0, the drift $g(x_0,\cdot)$ is estimated on a grid $(z_l)_{l=1,\dots,L} = (x_0, y_l)_{l=1,\dots,L}$.

For some fixed value of y_0, the drift $g(\cdot, y_0)$ is estimated on a grid $(z_l)_{l=1,\dots,L} = (x_l, y_0)_{l=1,\dots,L}$.

On Figure 7.1 below we chose $L = 40$ and $x_0 = 0.0230$.

On Figure 7.2 below we chose $L = 40$ and $y_0 = 0.1878$.

6.2. *Model II: generalized Duffing oscillator.* We consider the noisy Duffing oscillator known as Kramers oscillator. The system (1.1) can now be written as

$$\begin{aligned} dX_t &= Y_t dt \\ dY_t &= \sigma\, dW_t - (\kappa Y_t + \alpha X_t^3 - \beta X_t)dt \end{aligned} \tag{6.3}$$

with σ, κ, α and $\beta > 0$. The potential is then $V(x) = \alpha\frac{x^4}{4} - \beta\frac{x^2}{2}$. The invariant density is in that case

$$p_s(x,y) = \frac{\sqrt{\kappa}}{\sqrt{\pi}\sigma C} \exp\left(\frac{-2\kappa}{\sigma^2}\left(\frac{\alpha\, x^4}{4} - \frac{\beta\, x^2}{2} + \frac{y^2}{2}\right)\right),$$

with C the normalizing constant.

And the drift is defined by $g(x,y) = -(\kappa y + \alpha x^3 - \beta x)$.

Once more, we make use of the explicit Euler scheme to simulate an approximated discrete sampling. The choice for the parameters is $\sigma = 1$, $\kappa = \alpha = \beta = 1$.

We take $n = 10^4$, $h_n = 0.30$, $b_{1n} = b_{2n} = 0.30$, and the step for the explicit Euler scheme $\delta = \frac{1}{10}h/3$.

The drift is estimated for some fixed value of x_0, $g(x_0,\cdot)$, on a grid $(z_l)_{l=1,\dots,L} = (x_0, y_l)_{l=1,\dots,L}$. It is also estimated for some fixed value of y_0, $g(\cdot, y_0)$, on a grid $(z_l)_{l=1,\dots,L} = (x_l, y_0)_{l=1,\dots,L}$.

On Figure 7.3 below we chose $L = 40$ and $x_0 = 0.0230$. On Figure 7.4 below we chose $L = 40$ and $y_0 = 0.1878$.

6.3. *Model III: Van der Pol oscillator.* We consider the Van der Pol oscillator defined by

$$\begin{aligned} dX_t &= Y_t dt \\ dY_t &= \sigma\, dW_t - ((c_1 X_t^2 - c_2)Y_t + \omega_0^2 X_t)dt \end{aligned} \tag{6.4}$$

with σ, c_1, c_2 and $\omega_0^2 > 0$. In the following we choose $\sigma = c_1 = c_2 = \omega_0 = 1$. The drift is then defined by $g(x,y) = -((x^2-1)y + 4x)$. The invariant density is unknown in that case. However, we know that it is the solution of the corresponding Fokker-Planck equation

$$\frac{1}{2}\frac{\partial^2 p_s(x,y)}{\partial y^2} - y\frac{\partial p_s(x,y)}{\partial x} + c(x,y)p_s(x,y) + (c(x,y)y + V'(x))\frac{\partial p_s(x,y)}{\partial y} = 0\,. \quad (6.5)$$

Thus the invariant density $p_s(x,y)$ may be approximated by solving Equation (6.5) above, e.g. using a finite difference scheme. Doing so we remark that this density, although positive, is very small in many points. Therefore we plotted $g(x,y)\,p_s(x,y)$ to avoid numerical instabilities.

We made use of the explicit Euler scheme to simulate an approximated discrete sampling.

On Figures 7.5 and 7.6 below we took $n = 10^5$, $h_n = 0.18$, $b_{1n} = b_{2n} = 0.10$, and the step for the explicit Euler scheme $\delta = \frac{1}{10}h/3$.

The drift for some fixed value of x_0, $g(x_0,\cdot)$, is estimated on a grid $(z_l)_{l=1,\ldots,L} = (x_l, y_0)_{l=1,\ldots,L}$.

On Figure 7.5 below we chose $x_0 = 0.0230$ and plotted $g(x_0,\cdot) * p_s(x_0,\cdot)$.

The drift for some fixed value of y_0, $g(\cdot,y_0)$ is estimated on a grid $(z_l)_{l=1,\ldots,L} = (x_0, y_l)_{l=1,\ldots,L}$.

L was chosen equal to 40.

On Figure 7.6 below we chose $y_0 = 0.1878$ and plotted $g(\cdot,y_0) * p_s(\cdot,y_0)$.

We remark that for this third model, we do not estimate the drift itself, but the drift multiplied by the invariant density. Indeed, the direct estimation of the drift required divisions by quantities close to zero. Even with a regularization term, the results were not convincing.

Remark 6.1. In the simulations above we used the explicit Euler scheme, which may be unstable because the coefficients of the differential system 1.1 are unbounded for the models we considered as; see e.g. Talay (2002). We also implemented an implicit scheme, but it slowed too much the simulations, and the results were not better.

To conclude this section, note that non parametric estimation allows capturing the shape of a drift for which no parametric formula is available. For that reason, it is of real interest for practicians. Once the shape is captured, it might then become possible to propose a parametric model whose parameters should be estimated.

7. Appendix

7.1. *Proof of Proposition 4.2.* We may replace $\sqrt{n}$ by $\sqrt{n-1}$ without any change. Now decompose

$$\begin{aligned}
\mathcal{S}_n \quad &:= \quad \sqrt{(n-1)b_{1n}^d b_{2n}^d h_n}\big(\tilde{H}_n(x,y) - g(x,y)p_s(x,y)\big)\\
&= \quad \sqrt{(n-1)b_{1n}^d b_{2n}^d h_n}\big(\tilde{H}_n(x,y) - \mathbb{E}\tilde{H}_n(x,y) + \mathbb{E}\tilde{H}_n(x,y) - g(x,y)p_s(x,y)\big)\\
&=: \quad \mathcal{I}_{1n} + \mathcal{I}_{2n}.
\end{aligned}$$

To prove Proposition 4.2 we first prove that $\mathcal{I}_{2n} \to 0$ and then that

$$\mathcal{I}_{1n} \xrightarrow[n\to+\infty]{\mathcal{D}} \mathcal{N}(0, (\frac{2}{3}\,\sigma^2 p_s(x,y)\int K^2(x,y)dxdy)\mathcal{I}).$$

Define

$$I_i(h_n) := 9\left(\int_{(i+\frac{2}{3})h_n}^{(i+1)h_n}\int_{(i+\frac{2}{3})h_n}^{t} g(X_s,Y_s)dsdt + \int_{(i+\frac{1}{3})h_n}^{(i+\frac{2}{3})h_n}\int_{t}^{(i+\frac{2}{3})h_n} g(X_s,Y_s)dsdt\right),$$

and

$$\mathfrak{W}_{i,n} := 9\left(\int_{(i+\frac{2}{3})h_n}^{(i+1)h_n}(W_s - W_{(i+\frac{2}{3})h_n})ds + \int_{(i+\frac{1}{3})h_n}^{(i+\frac{2}{3})h_n}(W_{(i+\frac{2}{3})h_n} - W_s)ds\right).$$

The vector $\sqrt{(n-1)b_{1n}^d b_{2n}^d h_n}\tilde{H}_n(x,y)$ can be decomposed in two terms: the one driving the bias in the central limit theorem

$$S_{n,1}(x,y) := \frac{\sqrt{h_n}}{\sqrt{(n-1)b_{1n}^d b_{2n}^d}}\sum_{i=1}^{n-1} K\left(\frac{x-X_{ih_n}}{b_{1n}}, \frac{y-Y_{ih_n}}{b_{2n}}\right)\frac{1}{h_n^2} I_i(h_n),$$

and the one driving the variance

$$S_{n,2}(x,y) := \frac{\sigma}{\sqrt{(n-1)b_{1n}^d b_{2n}^d}}\sum_{i=1}^{n-1} K\left(\frac{x-X_{ih_n}}{b_{1n}}, \frac{y-Y_{ih_n}}{b_{2n}}\right)\frac{\mathfrak{W}_{i,n}}{h_n^{3/2}}.$$

Notice that $\mathbb{E}S_{n,2}(x,y) = 0$. We thus have

$$\mathcal{I}_{2n} = \mathbb{E}S_{n,1}(x,y) - \sqrt{(n-1)b_{1n}^d b_{2n}^d h_n}\, g(x,y)p_s(x,y), \tag{7.1}$$

while

$$\mathcal{I}_{1n} = (S_{n,1} - \mathbb{E}S_{n,1}(x,y)) + S_{n,2}(x,y). \tag{7.2}$$

First step: Study of $\mathcal{I}_{2n}$

We define

$$\frac{\mathfrak{P}_{i,n}}{9} := \int_{(i+\frac{2}{3})h_n}^{(i+1)h_n}\int_{(i+\frac{2}{3})h_n}^{t}(P_s g(X_{ih_n},Y_{ih_n}) - g(X_{ih_n},Y_{ih_n}))dsdt$$
$$+ \int_{(i+\frac{1}{3})h_n}^{(i+\frac{2}{3})h_n}\int_{t}^{(i+\frac{2}{3})h_n}(P_s g(X_{ih_n},Y_{ih_n}) - g(X_{ih_n},Y_{ih_n}))dsdt.$$

Thanks to stationarity, it holds

$$\mathcal{I}_{2n} = \sqrt{\frac{(n-1)h_n}{b_{1n}^d b_{2n}^d}}\,\mathbb{E}\left(K\left(\frac{x-X_0}{b_{1n}}, \frac{y-Y_0}{b_{2n}}\right)\frac{1}{h_n^2}\mathfrak{P}_{0,n}\right) +$$
$$+\sqrt{\frac{(n-1)h_n}{b_{1n}^d b_{2n}^d}}\left(\mathbb{E}\left(K\left(\frac{x-X_0}{b_{1n}}, \frac{y-Y_0}{b_{2n}}\right)g(X_0,Y_0)\right) - b_{1n}^d b_{2n}^d g(x,y)p_s(x,y)\right).$$

The second summand in the above expression can be treated as in a classical density estimation problem. More precisely, this term is equal to

$$\sqrt{(n-1)b_{1n}^d b_{2n}^d h_n}\int K(u,v)\left\{g(x-ub_{1n}, y-vb_{2n})p_s(x-ub_{1n}, y, vb_{2n}) - g(x,y)p_s(x,y)\right\}dudv.$$

Thus assuming there exists $m \in \mathbb{N}^*$ such that for all polynomials $P(x,y)$ with degree between 1 and m, $\int P(u,v)K(u,v)dudv = 0$, and performing a Taylor expansion, the above term converges to zero as n tends to infinity as soon as

$$n b_{1n}^d b_{2n}^d h_n \max(b_{1n}, b_{2n})^{2(m+1)} \xrightarrow[n\to+\infty]{} 0\,.$$

Let us now study the first summand. As each of the coordinates of the drift function g belongs to the domain of the infinitesimal generator L according to $\mathcal{H}_3$, $\forall\, 1 \le p < +\infty$ $(P_t g - g)/t$ is bounded in $\mathbb{L}^p(\mu)$ uniformly in t for $t \in [0,1]$, say by M_p.
Now write

$$\sqrt{\frac{(n-1)h_n}{b_{1n}^d b_{2n}^d}}\; \mathbb{E}\left(K\left(\frac{x-X_0}{b_{1n}}, \frac{y-Y_0}{b_{2n}}\right)\frac{1}{h_n^2}\mathfrak{P}_{0,n}\right) =: 9\sqrt{\frac{(n-1)h_n}{b_{1n}^d b_{2n}^d}}\;(A_{1n}+A_{2n})$$

with

$$A_{1n} = \frac{1}{h_n^2}\int\int_{2h_n/3}^{h_n}\int_{2h_n/3}^{t}(P_s g(u,v) - g(u,v))\; K\left(\frac{x-u}{b_{1n}}, \frac{y-v}{b_{2n}}\right) ds\, dt\, \mu(du,dv)$$

and A_{2n} being similar just changing $\int_{2h_n/3}^{h_n}\int_{2h_n/3}^{t}$ into $\int_{h_n/3}^{2h_n/3}\int_{t}^{2h_n/3}$. We thus only study A_{1n}.
Using Fubini's theorem we may first integrate with respect to t and write

$$A_{1n} = \frac{1}{h_n^2}\int\int_{2h_n/3}^{h_n}(h_n - s)\, s\left(\frac{P_s g(u,v) - g(u,v)}{s}\right) K\left(\frac{x-u}{b_{1n}}, \frac{y-v}{b_{2n}}\right) ds\, \mu(du,dv)\,.$$

Now we integrate with respect to μ, use Cauchy-Schwarz inequality and the remark we have made about $P_s g - g$ (assuming that $h_n \le 1$). We thus have ($|.|$ denoting the norm in $\mathbb{R}^d$)

$$\begin{aligned}|A_{1n}| &\le \frac{M_p}{h_n^2}\int_{2h_n/3}^{h_n}(h_n - s)\, s\left(\int K^r\left(\frac{x-u}{b_{1n}}, \frac{y-v}{b_{2n}}\right)\mu(du,dv)\right)^{1/r} ds\\ &\le C M_p\, h_n\left(\int K^r\left(\frac{x-u}{b_{1n}}, \frac{y-v}{b_{2n}}\right)\mu(du,dv)\right)^{1/r}\end{aligned}$$

with $p > 1\,,\ r < +\infty \in \mathbb{N}^*$ such that $\frac{1}{p} + \frac{1}{r} = 1$. It follows, using again the change of variables

$$\sqrt{\frac{(n-1)h_n}{b_{1n}^d b_{2n}^d}}\, A_{1n} \le C_p\,\sqrt{(n-1)h_n^3}\,(b_{1n}b_{2n})^{d(\frac{1}{r}-\frac{1}{2})}\;.$$

Thanks to assumption iii), one can choose r such that this last term tends to zero as n tends to infinity.

Second step: Study of $S_{n,2}$

We now consider the term driving the variance. To study the weak convergence of this sequence we adapt the proof of Theorem 3 in Beśka et al. (1982) and study the characteristic function of $S_{n,2}$. Let us recall that

$$S_{n,2}(x,y) := \frac{\sigma}{\sqrt{(n-1)b_{1n}^d b_{2n}^d}}\sum_{i=1}^{n-1} K\left(\frac{x - X_{ih_n}}{b_{1n}}, \frac{y - Y_{ih_n}}{b_{2n}}\right)\frac{\mathfrak{W}_{i,n}}{h_n^{3/2}}$$

and that $\mathbb{E}S_{n,2}(x,y)=0$.

The sketch of the proof of this step is the following. We first prove that the sequence of random variables $\{S_{n,2}(x,y)\,,\ n\geq 1\}$ is tight. We then prove that if $S_{n_k,2}(x,y)$ is a subsequence of the original sequence, which converges in distribution, then it converges to $Y\sim\mathcal{N}(0,1)$. We then conclude that the sequence itself converges in distribution to Y.

We proceed now with the proof.

Define for any $u\in\mathbb{R}_+$, $\mathcal{F}_u:=\sigma\left((X_l,Y_l)\,,\ 0\leq l\leq u\right)$. We now introduce, for $t\in\mathbb{R}^d$,

$$f_n(t):=\prod_{k=1}^{n-1}\mathbb{E}[e^{i<t,\frac{\sigma}{\sqrt{(n-1)b_{1n}^d b_{2n}^d}}K(\frac{x-X_{kh_n}}{b_{1n}},\frac{y-Y_{kh_n}}{b_{2n}})\frac{\mathfrak{W}_{k,n}}{h_n^{3/2}}>}|\mathcal{F}_{kh_n}]\,,$$

where $<\cdot,\cdot>$ denotes the usual scalar product in $\mathbb{R}^d$.

Thanks to the independence of the Brownian increments we get

$$f_n(t)=\prod_{k=1}^{n-1}e^{-\frac{t^2\frac{2}{3}\sigma^2 d}{2(n-1)b_{1n}^d b_{2n}^d}K^2(\frac{x-X_{kh_n}}{b_{1n}},\frac{x-Y_{kh_n}}{b_{2n}})}=e^{-\frac{t^2\frac{2}{3}\sigma^2 d}{2(n-1)b_{1n}^d b_{2n}^d}\sum_{k=1}^{n-1}K^2(\frac{x-X_{kh_n}}{b_{1n}},\frac{x-Y_{kh_n}}{b_{2n}})}.$$

Now define

$$Z_n=\frac{1}{(n-1)b_{1n}^d b_{2n}^d}\sum_{i=1}^{n-1}K^2\left(\frac{x-X_{ih_n}}{b_{1n}},\frac{y-Y_{ih_n}}{b_{2n}}\right).$$

It satisfies

$$\mathbb{E}Z_n=\frac{1}{b_{1n}^d b_{2n}^d}\int_{\mathbb{R}^{2d}}K^2\left(\frac{x-u}{b_{1n}},\frac{y-v}{b_{2n}}\right)p_s(u,v)dudv\to p_s(x,y)\int K^2(u,v)dudv=A.$$

Furthermore

$$\begin{aligned}\mathbb{E}\left(|Z_n-A|\right)&\leq\int\left|\left(\frac{1}{b_{1n}^d b_{2n}^d}K^2\left(\frac{x-u}{b_{1n}},\frac{y-v}{b_{2n}}\right)p_s(u,v)\right)-A\right|dudv\\&\leq\frac{1}{b_{1n}^d b_{2n}^d}\int K^2\left(\frac{x-u}{b_{1n}},\frac{y-v}{b_{2n}}\right)|p_s(u,v)-p_s(x,y)|\,dudv\\&\leq\int K^2(u,v)\,|p_s(x-b_{1n}u,y-b_{2n}v)-p_s(x,y)|\,dudv\end{aligned}$$

and the latter goes to 0 by using the bounded convergence theorem and the continuity of p_s.

Thus $Z_n\to p_s(x,y)\int K^2(z_1,z_2)dz_1dz_2$, in L^1. Using the bounded convergence theorem, we deduce that

$$f_n(t)\xrightarrow{\mathbb{P}}e^{-\frac{t^2\frac{2}{3}\sigma^2 d}{2}p_s(x,y)\int K^2(z_1,z_2)dz_1dz_2}=:\phi(t).$$

Passing to subsequences if necessary, we can assume that this convergence holds almost everywhere i.e.

$$f_n(t)\to\phi(t)\,a.e. \tag{7.3}$$

Let us now define for $k=1,\ldots,n$ the sets

$$H_{nk}=\{e^{-\frac{t^2\frac{2}{3}\sigma^2 d}{2(n-1)b_{1n}^d b_{2n}^d}\sum_{i=1}^{k-1}K^2(\frac{x-X_{ih_n}}{b_{1n}},\frac{y-Y_{ih_n}}{b_{2n}})}\geq\frac{1}{2}\phi(t)\}.$$

Of course

$$H_{n,n} \subset H_{n,n-1} \subset \ldots \subset H_{n,2},$$

and using (7.3)

$$\mathbb{P}\{\limsup_{n\to\infty} H^c_{n,n}\} = 0. \tag{7.4}$$

Finally, introduce the random variables

$$\zeta_{n,i} = \frac{\sigma}{\sqrt{(n-1)b^d_{1n}b^d_{2n}}} K\left(\frac{x - X_{ih_n}}{b_{1n}}, \frac{y - Y_{ih_n}}{b_{2n}}\right) \frac{\mathfrak{W}_{i,n}}{h_n^{3/2}} \mathbf{1}_{H_{n,i}},$$

and also

$$f^*_n(t) = \prod_{i=1}^{n-1} E[e^{i<t,\zeta_{n,i}>}|\mathcal{F}_{ih_n}] = \prod_{i=1}^{n-1} e^{-\frac{t^2\frac{2}{3}\sigma^2 d}{2(n-1)b^d_{1n}b^d_{2n}} K^2(\frac{x-X_{ih_n}}{b_{1n}}, \frac{y-Y_{ih_n}}{b_{2n}})\mathbf{1}_{H_{n,i}}}.$$

It holds

$$f^*_n(t) \geq \frac{1}{2}\phi(t) \quad \text{by definition and } f^*_n(t) \to \phi(t)\; a.e. \text{ by (7.4)}\ . \tag{7.5}$$

So, we can assume that these two properties hold for the initial variables.

Let us now come back to the study of the weak convergence of $S_{n,2}(x,y)$. By using Markov property and the independence of the Brownian increments, we obtain

$$\mathbb{E}\left[\frac{e^{i<t,S_{n,2}(x,y)>}}{f_n(t)}\right] = E\left[\prod_{i=1}^{n-1} \frac{e^{i<t, \frac{\sigma}{\sqrt{(n-1)b^d_{1n}b^d_{2n}}} K(\frac{x-X_{ih_n}}{b_{1n}}, \frac{y-Y_{ih_n}}{b_{2n}}) \frac{\mathfrak{W}_{i,n}}{h_n^{3/2}}>}}{e^{-\frac{t^2\frac{2}{3}\sigma^2 d}{2(n-1)b^d_{1n}b^d_{2n}} K^2(\frac{x-X_{ih_n}}{b_{1n}}, \frac{x-Y_{ih_n}}{b_{2n}})}}\right]$$

$$= \mathbb{E}\left[\prod_{i=1}^{n-2} \frac{e^{i<t, \frac{\sigma}{\sqrt{(n-1)b^d_{1n}b^d_{2n}}} K(\frac{x-X_{ih_n}}{b_{1n}}, \frac{y-Y_{ih_n}}{b_{2n}}) \frac{\mathfrak{W}_{i,n}}{h_n^{3/2}}>}}{e^{-\frac{t^2\frac{2}{3}\sigma^2 d}{2(n-1)b^d_{1n}b^d_{2n}} K^2(\frac{x-X_{ih_n}}{b_{1n}}, \frac{x-Y_{ih_n}}{b_{2n}})}} \mathbb{E}\left[\frac{e^{i<t, \frac{\sigma}{\sqrt{(n-1)b^d_{1n}b^d_{2n}}} K(\frac{x-X_{(n-1)h_n}}{b_{1n}}, \frac{y-Y_{(n-1)h_n}}{b_{2n}}) \frac{\mathfrak{W}_{(n-1),n}}{h_n^{3/2}}>}}{e^{-\frac{t^2\frac{2}{3}\sigma^2 d}{2(n-1)b^d_{1n}b^d_{2n}} K^2(\frac{x-X_{(n-1)h_n}}{b_{1n}}, \frac{x-Y_{(n-1)h_n}}{b_{2n}})}} \middle| \mathcal{F}_{(n-1)h}\right]\right]$$

$$= \mathbb{E}\left[\prod_{i=1}^{n-2} \frac{e^{i<t, \frac{\sigma}{\sqrt{(n-1)b^d_{1n}b^d_{2n}}} K(\frac{x-X_{ih_n}}{b_{1n}}, \frac{y-Y_{ih_n}}{b_{2n}}) \frac{\mathfrak{W}_{i,n}}{h_n^{3/2}}>}}{e^{-\frac{t^2\frac{2}{3}\sigma^2 d}{2(n-1)b^d_{1n}b^d_{2n}} K^2(\frac{x-X_{ih_n}}{b_{1n}}, \frac{x-Y_{ih_n}}{b_{2n}})}}\right] = 1 \text{ using induction.}$$

Now we are ready to prove the weak convergence

$$\left|\mathbb{E}\left[e^{i<t,S_{n,2}(x,y)>}\right] - \phi(t)\right| = \left|\mathbb{E}\left[e^{i<t,S_{n,2}(x,y)>} - \mathbb{E}\left[\phi(t)\frac{e^{i<t,S_{n,2}(x,y)>}}{f_n(t)}\right]\right]\right|$$

$$= \left|\mathbb{E}\left[e^{i<t,S_{n,2}(x,y)>}\left(1 - \frac{\phi(t)}{f_n(t)}\right)\right]\right| \leq \frac{2}{\phi(t)}\mathbb{E}|f_n(t) - \phi(t)| \to 0.$$

This last equality is deduced from the $\mathbb{L}^1$-$\mathbb{L}^\infty$ Hölder inequality combined with the bound on $f_n(t)$ deduced from (7.5).

We may conclude the proof of the tightness of $\{S_{n,2}(x,y)\,,\ n \geq 1\}$. To this end, let $S_{n_k,2}(x,y)$ be a subsequence of the original sequence. We know that $f_{n_k}(t) \overset{\mathbb{P}}{\to} \phi(t)$. Whence there exists another subsequence $f_{n_{k_j}}(t) \overset{\text{a.e.}}{\to} \phi(t)$. By the above result $S_{n_{k_j}}$ converges weakly to a r.v. Y, moreover $\mathbb{E}[e^{i<t,Y>}] = \phi(t)$. Thus the tightness.

All the limits of the convergent subsequences being the same, we directly conclude that the sequence $\{S_{n,2}(x,y)\,,\ n \geq 1\}$ converges weakly and that its limit is Y. It concludes the proof of the second step.

Third step: study of $\mathcal{Z}_n := S_{n,1}(x,y) - \mathbb{E}S_{n,1}(x,y)$

Let us denote as before by $\mathfrak{P}^k_{i,n}$ the k^{th} coordinate of the vector $\mathfrak{P}_{i,n}$. Defining
$\Gamma^k_n(i,x,y,X,Y) = K(\frac{x-X}{b_{1n}}, \frac{y-Y}{b_{2n}})\frac{1}{h_n^2}\mathfrak{P}^k_{i,n}$, we write

$$\mathcal{Z}^k_n = \frac{\sqrt{h_n}}{\sqrt{(n-1)b_{1n}^d b_{2n}^d}} \sum_{i=1}^{n-1} \Gamma^k_n(i,x,y,X_{ih_n},Y_{ih_n}),$$

so that

$$\frac{(n-1)b_{1n}^d b_{2n}^d}{h_n} Var(\mathcal{Z}^k_n)$$
$$= \Big(\sum_{i=1}^{n-1} Var(\Gamma^k_n(i,x,y,X_{ih_n},Y_{ih_n})) + \sum_{i\neq l} Cov(\Gamma^k_n(i,x,y,X_{ih_n},Y_{ih_n}),\Gamma^k_n(l,x,y,X_{lh_n},Y_{lh_n}))\Big)\,.$$

To bound the above expression we first write as we did for the first step

$$\mathbb{E}(\Gamma^k_n(i,x,y,X_{ih_n},Y_{ih_n}))^2 = \mathbb{E}\left(K(\frac{x-X}{b_{1n}}, \frac{y-Y}{b_{2n}})\frac{1}{h_n^2}\mathfrak{P}^k_{i,n}\right)^2 \leq U_{1n} + U_{2n}$$

with

$$U_{1n} = 2\left(9\int_{(i+\frac{2}{3})h_n}^{(i+1)h_n}\int_{(i+\frac{2}{3})h_n}^{t} (P_s g(X_{ih_n},Y_{ih_n}) - g(X_{ih_n},Y_{ih_n}))dsdt\right)^2$$

and

$$U_{2n} = 2\left(9\int_{(i+\frac{1}{3})h_n}^{(i+\frac{2}{3})h_n}\int_{t}^{(i+\frac{2}{3})h_n} (P_s g(X_{ih_n},Y_{ih_n}) - g(X_{ih_n},Y_{ih_n}))dsdt\right)^2.$$

Then, using stationarity,

$$U_{1n} \leq \frac{C}{h_n^4}\,\mathbb{E}\left[K^2\left(\frac{x-X_0}{b_{1n}}, \frac{y-Y_0}{b_{2n}}\right)\left(\int_{2h_n/3}^{h_n}(h_n - s)\left(P_s g(X_0,Y_0) - g(X_0,Y_0)\right)\,ds\right)^2\right],$$

U_{2n} being similar just replacing $\int_{2h_n/3}^{h_n}(h_n - s)$ by $\int_{h_n/3}^{2h_n/3}(s-(h_n/3))$.
Using Cauchy-Schwarz inequality we get

$$\begin{aligned} U_{1n} &\leq \frac{C}{h_n}\int_{2h_n/3}^{h_n}\left(\int K^2\left(\frac{x-u}{b_{1n}}, \frac{y-v}{b_{2n}}\right)\left(P_s g(u,v) - g(u,v)\right)^2\,d\mu\right)ds \\ &\leq C\,h_n\int_{2h_n/3}^{h_n}\left(\int K^2\left(\frac{x-u}{b_{1n}}, \frac{y-v}{b_{2n}}\right)\left(\frac{P_s g(u,v) - g(u,v)}{s}\right)^2\,d\mu\right)ds\,. \end{aligned}$$

We may argue as in the first step, this time using Hölder inequality for some conjugate pair (p,q) and $\mathcal{H}_3$ in $\mathbb{L}^{2q}$, to conclude that

$$U_{1n} \leq C\,h_n^2\,(b_{1n}b_{2n})^{d/p}\,.$$

It follows

$$\frac{h_n}{(n-1)b_{1n}^d b_{2n}^d}\sum_{i=1}^{n-1} Var(\Gamma_n^k(i,x,y,X_{ih_n},Y_{ih_n})) = \mathcal{O}\left(h_n^3 (b_{1n}b_{2n})^{d(\frac{1}{p}-1)}\right). \tag{7.6}$$

One can choose p such that the right hand term tends to zero thanks to assumption iv).

Let us now compute the covariances.
One has thanks to stationarity and mixing inequality (2.1)

$$\sum_{i\neq l} \mathrm{Cov}\left(\Gamma_n^k(i,x,y,X_{ih_n},Y_{ih_n}),\Gamma_n^k(l,x,y,X_{lh_n},Y_{lh_n})\right)$$

$$\leq Cn\sum_{j=1}^{n-2}\min\left(\rho^{jh_n/2},\mathrm{Var}(\Gamma_n^k(0,x,y,X_0,Y_0))\right).$$

Then, using inequality (7.6) one gets the following bound

$$\leq Cn\sum_{j=1}^{n-2}\min\left(\rho^{jh_n/2},(b_{1n}b_{2n})^{\frac{d}{p}}h_n^2\right).$$

Then using that for all $x,y\geq 0$, and all $a\in(0,1)$, $x\wedge y\leq x^{1-a}y^a$ one gets that the above quantity is in

$$\mathcal{O}\left(n(b_{1n}b_{2n})^{\frac{d}{p}(1-a)}h_n^{1-2a}\right)\quad \text{for any } 0<a<1.$$

We thus get

$$\frac{h_n}{(n-1)b_{1n}^d b_{2n}^d}\sum_{i\neq l} \mathrm{Cov}\left(\Gamma_n^k(i,x,y,X_{ih_n},Y_{ih_n}),\Gamma_n^k(l,x,y,X_{lh_n},Y_{lh_n})\right)$$

$$=\mathcal{O}\left((b_{1n}b_{2n})^{\frac{d}{p}(1-a)-d}h_n^{2(1-a)}\right).$$

One can choose p and a such that the right hand term tends to zero as n tends to infinity thanks to assumption iv). This completes the proof.

7.2. *Proof of Proposition 4.3*. Starting from Proposition 4.2, it remains now to consider D_n defined by

$$\frac{1}{\sqrt{(n-1)b_{1n}^d b_{2n}^d}}\sum_{j=1}^{n-1}\left[K\left(\frac{x-X_{jh_n}}{b_{1n}},\frac{y-Y_{jh_n}}{b_{2n}}\right)-K\left(\frac{x-X_{jh_n}}{b_{1n}},\frac{y-\frac{X_{(j+\frac{1}{3})h_n}-X_{jh}}{(h_n/3)}}{b_{2n}}\right)\right]\frac{\mathfrak{D}_{j,n}}{h^{3/2}}.$$

Let us define $A_j = K\left(\frac{x-X_{jh_n}}{b_{1n}},\frac{y-Y_{jh_n}}{b_{2n}}\right)-K\left(\frac{x-X_{jh_n}}{b_{1n}},\frac{y-\frac{X_{(j+\frac{1}{3})h_n}-X_{jh}}{(h_n/3)}}{b_{2n}}\right)$. We then write

$$D_n=\frac{1}{\sqrt{(n-1)b_{1n}^d b_{2n}^d}}\sum_{j=1}^{n-1}A_j\,\frac{\mathfrak{D}_{j,n}}{h^{3/2}}=\frac{1}{\sqrt{(n-1)b_{1n}^d b_{2n}^d}}\sum_{j=1}^{n-1}A_j\,\frac{(\sigma\,\mathfrak{W}_{j,n}+I_j(h_n))}{h^{3/2}}.$$

Using Hölder Inequality, we bound $\mathbb{E}\left|\frac{1}{\sqrt{(n-1)b_{1n}^d b_{2n}^d}}\sum_{j=1}^{n-1} A_j\,\frac{I_j(h_n)}{h^{3/2}}\right|$ by

$$\frac{1}{\sqrt{(n-1)b_{1n}^d b_{2n}^d}}h^{-3/2}\sum_{j=1}^{n-1}\left(\mathbb{E}|A_j|^{1+\varepsilon}\right)^{1/(1+\varepsilon)}\left(\mathbb{E}|I_j(h_n)|^{(1+\varepsilon)/\varepsilon}\right)^{\varepsilon/(1+\varepsilon)},$$

with $\varepsilon > 0$.

We first consider $\mathbb{E}|I_j(h_n)|^{(1+\varepsilon)/\varepsilon} = \mathbb{E}|I_0(h_n)|^{(1+\varepsilon)/\varepsilon}$ by stationarity. We use (4.4) and we get that

$$\mathbb{E}\left[|I_0(h_n)|^{(1+\varepsilon)/\varepsilon}|X_0=u, Y_0=v\right] \le \left|\int_{\frac{2}{3}h_n}^{h_n}\int_{\frac{2}{3}h_n}^{t} P_s g(u,v)dsdt + \int_{\frac{h_n}{3}}^{\frac{2}{3}h_n}\int_{t}^{\frac{2}{3}h_n} P_s g(u,v)dsdt\right|^{(1+\varepsilon)/\varepsilon}. \tag{7.7}$$

Recall that for all $t>0$, and for all $f\in\mathbb{L}^p$, $1\le p\le\infty$, $P_t f(z)$ is defined as $\mathbb{E}_z f(Z_t)$.

Thus for all $t>0$ and all $\alpha>0$

$$|P_t g(u,v)|^{\alpha} = \left|\int g(x,y)p_t(x,y;u,v)dxdy\right|^{\alpha} \le \int |g(x,y)|^{\alpha} p_t(x,y;u,v)dxdy = P_t|g|^{\alpha}(u,v)$$

with $p_t(z;z')$ denoting the transition kernel for the Markov process $\mathbf{Z}$.

We can thus bound the right hand side of (7.7), using $\mathbb{L}^{\frac{1+\varepsilon}{\varepsilon}}$-$\mathbb{L}^{1+\varepsilon}$ Hölder inequality, by

$$C\,h_n^{\frac{2}{\varepsilon}}\left[\int_{\frac{2}{3}h_n}^{h_n}\int_{\frac{2}{3}h_n}^{t} P_s|g|^{\frac{1+\varepsilon}{\varepsilon}}(u,v)dsdt + \int_{\frac{h_n}{3}}^{\frac{2}{3}h_n}\int_{t}^{\frac{2}{3}h_n} P_s|g|^{\frac{1+\varepsilon}{\varepsilon}}(u,v)dsdt\right]$$

which is equal to

$$C\,h_n^{\frac{2}{\varepsilon}}h_n^2 G_\varepsilon(h_n,u,v)$$

with $G_\varepsilon(h_n,u,v)\xrightarrow[n\to+\infty]{}|g|^{\frac{1+\varepsilon}{\varepsilon}}(u,v)$.

Using now similar arguments as in Step 2 in the proof of Theorem 3.3 in Cattiaux et al. (2014) (Theorem 3.1 of the present paper), one bounds $\mathbb{E}|A_i|^{1+\varepsilon}$ by

$$\mathcal{O}\left(h_n^{1+\varepsilon}b_{1n}^{\frac{d}{p}}b_{2n}^{-(1+\varepsilon)}\right). \tag{7.8}$$

Thus $\mathbb{E}\left|\frac{1}{\sqrt{(n-1)b_{1n}^d b_{2n}^d}}\sum_{j=1}^{n-1} A_j\,\frac{I_j(h_n)}{h^{3/2}}\right|$ is bounded by

$$\mathcal{O}\left(h_n\sqrt{n}\,b_{1n}^{d(\frac{1}{p(1+\varepsilon)}-\frac{1}{2})}b_{2n}^{-(\frac{d}{2}+1)}\right)$$

with $1<p<+\infty$. It converges to zero as we assumed (see Assumption v).

It remains now to bound $\mathbb{E}\left|\frac{1}{\sqrt{(n-1)b_{1n}^d b_{2n}^d}}\sum_{j=1}^{n-1} A_j\,\frac{\sigma\,\mathfrak{W}_{j,n}}{h^{3/2}}\right|$. The terms $A_j\,\mathfrak{W}_{j,n}$, $1\le j\le n-1$ are centered and uncorrelated. Thus by stationarity we have to bound $\frac{\sigma^2}{b_{1n}^d\,b_{2n}^d}\mathbb{E}\frac{A_0^2\mathfrak{W}_{0,n}^2}{h_n^3}$. First conditioning now on $Z_{\frac{h_n}{3}}$ one gets $\mathbb{E}A_0^2\mathfrak{W}_{0,n}^2 = h_n^3\mathbb{E}A_0^2$. We then conclude by using (7.8) for $\varepsilon=1$ which yields $\mathbb{E}A_0^2\mathfrak{W}_{0,n}^2 = h_n^3 h_n^2 b_{1n}^{d/p}b_{2n}^{-2}$ for all $1<p<+\infty$. This completes the proof, using v).

7.3. *Proof os Proposition 4.4*. For Theorem 3.1 to be satisfied in this situation we must have (see Remark 3.4 in Cattiaux et al. (2014))

(a) $\beta_1 > \frac{2+\beta_2(3+2d)}{1+2d}$ in particular $\beta_1 > \beta_2$,
(b) $1 - \beta_1 d + \beta_2(2+d) < \gamma < \frac{1}{2}(\beta_1+\beta_2) < \frac{1}{2d}$,
(c) $m > \frac{1-d(\beta_1+\beta_2)}{2\beta_2}$.

Condition (a) ensures that $1 - \beta_1 d + \beta_2(2+d) < \frac{1}{2}(\beta_1+\beta_2)$, so that we may find some γ sandwiched by both terms. For both (a) and (b) to be satisfied, it is necessary that

$$\beta_2 < \frac{1}{d(4+4d)}\,, \tag{7.9}$$

and then we can choose

$$\frac{1}{d} - \beta_2 > \beta_1 > \frac{2+\beta_2(3+2d)}{1+2d}\,.$$

Remark that our interest is to take β_2 as close as possible to its upper bound, β_1 as close as possible to $\frac{1}{d} - \beta_2$ so that m can be chosen as small as possible (1 is possible). γ can then be chosen smaller than and close to $1/2d$.

Now we look at Proposition 4.3 starting with conditions i)-iv) in Proposition 4.2

(1) $1 > \gamma + d(\alpha_1+\alpha_2)$,
(2) $1 < d(\alpha_1+\alpha_2) + 2(m+1)(\alpha_1 \wedge \alpha_2) + \gamma$,
(3) $1 < 3\gamma + d(\alpha_1+\alpha_2)$ is enough for getting $\varepsilon_1 > 0$ in iii),
(4) $2\gamma > \varepsilon_3\,(\alpha_1+\alpha_2)$ is enough for getting $\varepsilon_2 > 0$ in iv).

We see that the latter will be automatically satisfied for ε_3 small enough. We have to add

(5) $2\gamma > (2+d)\alpha_2$,
(6) $\frac{1}{2} < \gamma + \frac{d}{2}\,(\alpha_1-\alpha_2) - \alpha_2$,

which is enough to furnish both $p > 1$ and $\varepsilon > 0$ in v) of Proposition 4.3.

Finally we have to add

$$(7) \quad \gamma > d((\beta_1+\beta_2) - (\alpha_1+\alpha_2))\,.$$

Look at (1). The compatibility with (b) imposes

$$\gamma + d(\alpha_1+\alpha_2) < 1 < \gamma + d(\beta_1+\beta_2) - \beta_2(2+2d)\,.$$

That is why we cannot take the same bandwidths b and c.
We thus have a first necessary condition

$$d(\alpha_1+\alpha_2) < d(\beta_1+\beta_2) - \beta_2(2+2d)\,,$$

which is satisfied as soon as

$$d(\alpha_1+\alpha_2) = d(\beta_1+\beta_2) - \frac{1}{2d}\,. \tag{7.10}$$

Now we choose for some small ε (say less than $10^{-6}/d^2$),

$$\beta_2 = \frac{1-\varepsilon}{d(4+4d)} \quad , \quad \beta_1 = \frac{3-2\varepsilon+4d}{d(4+4d)}\,.$$

(7.9) and the following inequalities are satisfied, as well as the final upper bound in (b). We have thus to choose γ such that, first $\gamma > \frac{1}{2d}$ for (7) to be satisfied, next for (1) and the first lower bound of (b) to be satisfied. This amounts to

$$\frac{1-\varepsilon}{2d} + \frac{3\varepsilon}{4+4d} < \gamma < \frac{1}{2d} + \frac{3\varepsilon}{4+4d} \,.$$

The left hand side of the previous inequality is less than $\frac{1}{2d}$ since $d \geq 1$. So we only have to choose

$$\frac{1}{2d} < \gamma < \frac{1}{2d} + \frac{3\varepsilon}{4+4d} \,. \tag{7.11}$$

The upper bound in (b) are then satisfied. If we look at (3) it reduces to

$$\frac{1}{2d} + \frac{3\varepsilon}{4+4d} < 3\gamma$$

which is satisfied. Now (6) becomes

$$\frac{3\varepsilon}{4+4d} + \frac{1}{2d} + (2d+2)\alpha_2 < 2\gamma$$

which implies (5) and which is satisfied as soon as

$$\frac{3\varepsilon}{4+4d} + (2d+2)\alpha_2 < \frac{1}{2d} \,.$$

Notice that the latter implies $\alpha_2 < \alpha_1$. Notice that for $\varepsilon > 0$ we may choose $\alpha_2 = \beta_2 - \frac{\varepsilon}{d(4+4d)}$. It remains to choose m for (2) to be satisfied.

Acknowledgements

The authors would like to thank both referees for their very careful reading of their work.

References

Dominique Bakry, Patrick Cattiaux and Arnaud Guillin. Rate of convergence for ergodic continuous Markov processes: Lyapunov versus Poincaré. *J. Funct. Anal.* **254** (3), 727–759 (2008).

Marek Beśka, Andrzej Kłopotowski and Leszek Słomiński. Limit theorems for random sums of dependent d-dimensional random vectors. *Z. Wahrsch. Verw. Gebiete* **61** (1), 43–57 (1982).

Patrick Cattiaux, José R. León and Clémentine Prieur. Estimation for stochastic damping hamiltonian systems under partial observation—I. Invariant density. *Stochastic Process. Appl.* **124** (3), 1236–1260 (2014).

Fabienne Comte, Valentine Genon-Catalot and Yves Rozenholc. Penalized nonparametric mean square estimation of the coefficients of diffusion processes. *Bernoulli* **13** (2), 514–543 (2007).

Arnak Dalalyan. Sharp adaptive estimation of the drift function for ergodic diffusions. *Ann. Statist.* **33** (6), 2507–2528 (2005).

C. W. Gardiner. *Handbook of stochastic methods*, volume 13 of *Springer Series in Synergetics*. Springer-Verlag, Berlin, second edition (1985). For physics, chemistry and the natural sciences.

Arnaud Gloter. Efficient estimation of drift parameters in stochastic volatility models. *Finance Stoch.* **11** (4), 495–519 (2007).

Yvo Pokern, Andrew M. Stuart and Petter Wiberg. Parameter estimation for partially observed hypoelliptic diffusions. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **71** (1), 49–73 (2009).

Adeline Samson and Michèle Thieullen. A contrast estimator for completely or partially observed hypoelliptic diffusion. *Stochastic Process. Appl.* **122** (7), 2521–2552 (2012).

D. Talay. Stochastic Hamiltonian systems: exponential convergence to the invariant measure, and discretization by the implicit Euler scheme. *Markov Process. Related Fields* **8** (2), 163–198 (2002). Inhomogeneous random systems (Cergy-Pontoise, 2001).

Liming Wu. Large and moderate deviations and exponential convergence for stochastic damping Hamiltonian systems. *Stochastic Process. Appl.* **91** (2), 205–238 (2001).

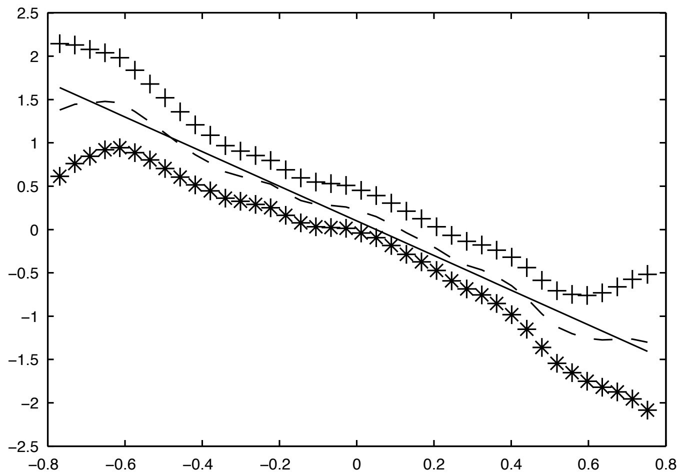

FIGURE 7.1. Estimation of the drift for the harmonic oscillator: theoretical $g(x_0, \cdot)$ in plain line, estimated in dashed line for $x_0 = 0.0230$, 95% asymptotic confidence intervals (upper bounds with croices, lower bounds with stars)

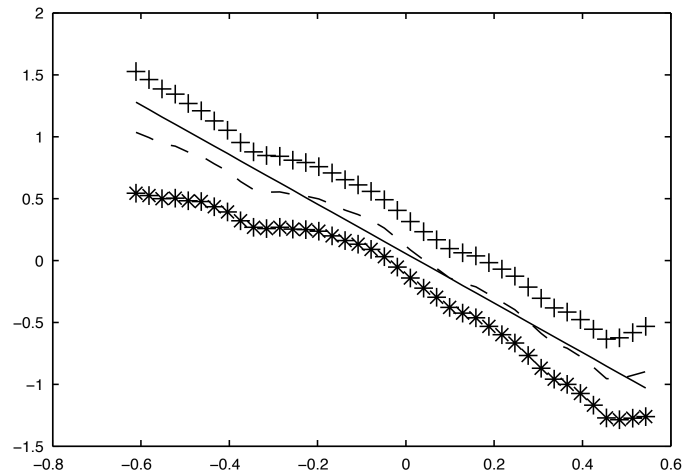

FIGURE 7.2. Estimation of the drift for the harmonic oscillator: theoretical $g(\cdot, y_0)$ in plain line, estimated in dashed line for $y_0 = 0.1878$, 95% asymptotic confidence intervals (upper bounds with croices, lower bounds with stars)

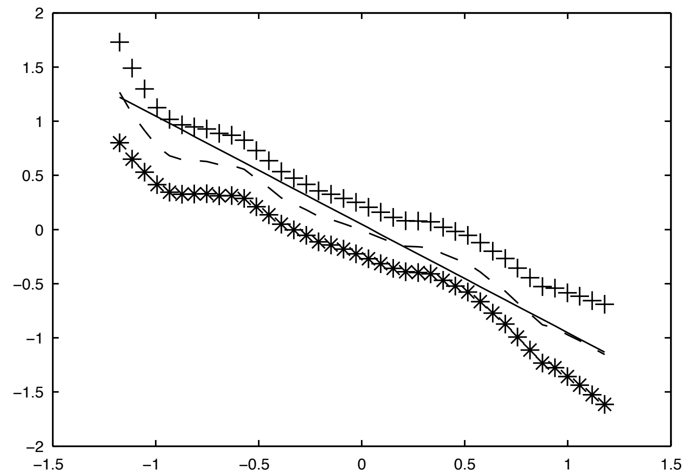

Figure 7.3. Estimation of the drift for the Duffing oscillator: theoretical $g(x_0, \cdot)$ in plain line, estimated in dashed line for $x_0 = 0.0230$, 95% asymptotic confidence intervals (upper bounds with croices, lower bounds with stars)

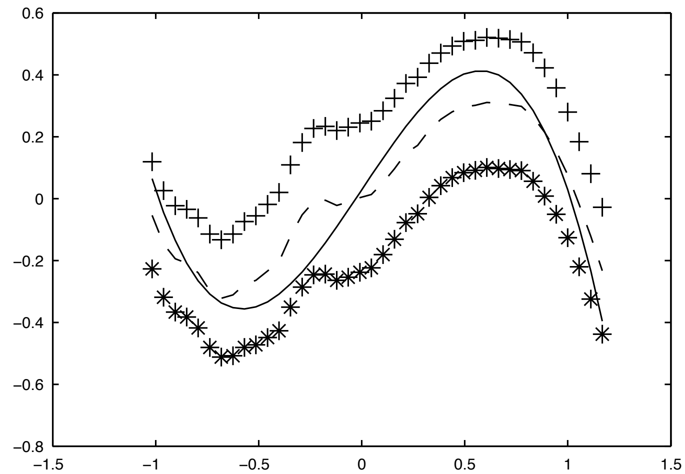

FIGURE 7.4. Estimation of the drift for the Duffing oscillator: theoretical $g(\cdot, y_0)$ in plain line, estimated in dashed line for $y_0 = 0.1878$, 95% asymptotic confidence intervals (upper bounds with croices, lower bounds with stars)

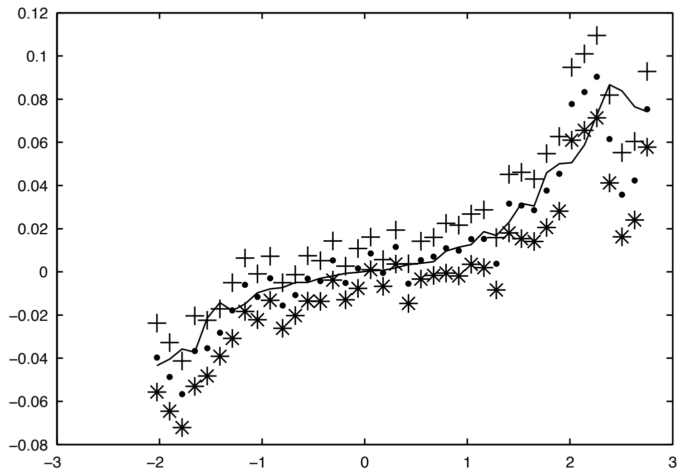

FIGURE 7.5. Estimation of the "drift" for the Van der Pol oscillator: theoretical $g(x_0, \cdot) * p_s(x_0, \cdot)$ in plain line, estimated with dots for $x_0 = 0.0230$, 95% asymptotic confidence intervals (upper bounds with croices, lower bounds with stars)

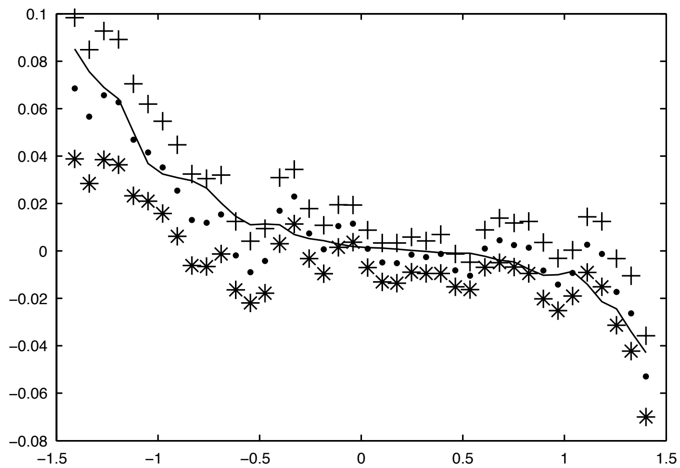

FIGURE 7.6. Estimation of the "drift" for the Van der Pol oscillator: theoretical $g(\cdot, y_0) * p_s(\cdot, y_0)$ in plain line, estimated with dots for $y_0 = 0.1878$, 95% asymptotic confidence intervals (upper bounds with croices, lower bounds with stars)